Symposium on Theoretical Aspects of Computer Science 2009 (Freiburg), pp. 601–612
www.stacs-conf.org

POLYNOMIAL KERNELIZATIONS FOR MIN $F^+\Pi_1$ AND MAX NP

STEFAN KRATSCH [1]

[1] Max-Planck-Institut für Informatik, Campus E 1 4, 66123 Saarbrücken, Germany
E-mail address: skratsch@mpi-inf.mpg.de
URL: http://www.mpi-inf.mpg.de/~skratsch

Abstract. The relation of constant-factor approximability to fixed-parameter tractability and kernelization is a long-standing open question. We prove that two large classes of constant-factor approximable problems, namely MIN $F^+\Pi_1$ and MAX NP, including the well-known subclass MAX SNP, admit polynomial kernelizations for their natural decision versions. This extends results of Cai and Chen (JCSS 1997), stating that the standard parameterizations of problems in MAX SNP and MIN $F^+\Pi_1$ are fixed-parameter tractable, and complements recent research on problems that do not admit polynomial kernelizations (Bodlaender et al. ICALP 2008).

1. Introduction

The class APX consists of all NP optimization problems that are approximable to within a constant factor of the optimum. It is known that the decision versions of most APX-problems are fixed-parameter tractable or even admit efficient preprocessing in the form of a polynomial kernelization. How strong is the relation between constant-factor approximability and polynomial kernelizability? Is there a property inherent to most APX-problems that explains this relation? What is the nature of APX-problems that do not admit a polynomial kernelization, such as Bin Packing for example?

Since many prominent APX-problems are complete under approximation preserving reductions and do not admit arbitrarily small approximation ratios, studying their parameterized complexity is a natural approach to obtain better results (recently Cai and Huang presented fixed-parameter approximation schemes for MAX SNP [7]). In conjunction with recent work on problems without polynomial kernelizations, positive answers to the questions may provide evidence against APX-membership for some problems (e.g. Treewidth).

Our work: We prove that the standard parameterizations of problems in two large classes of constant-factor approximable problems, namely MIN $F^+\Pi_1$ and MAX NP, admit polynomial kernelizations. This extends results of Cai and Chen [6] who showed that the standard parameterizations of all problems in MIN $F^+\Pi_1$ and MAX SNP (a subclass of MAX NP) are fixed-parameter tractable.[1] Interestingly perhaps, both our results rely on the Sunflower Lemma due to Erdős and Rado [10].

1998 ACM Subject Classification: F.2.2.
Key words and phrases: parameterized complexity, kernelization, approximation algorithms.

[1] The existence of a kernelization, not necessarily polynomial, is equivalent to fixed-parameter tractability.

	Approximation ratio	Kernel size
Minimum Vertex Cover	2 [15]	$O(k)$ [8]
Feedback Vertex Set	2 [3]	$O(k^3)$ [4]
Minimum Fill-In	$O(\text{opt})$ [19]	$O(k^2)$ [19]
Treewidth	$O(\sqrt{\log \text{opt}})$ [12]	not poly[2] [5]

Table 1: Approximation ratio and size of problem kernels for some optimization problems.

Related work: Recently Bodlaender et al. [5] presented the first negative results concerning the existence of polynomial kernelizations for some natural fixed-parameter tractable problems. Using the notion of a *distillation algorithm* and results due to Fortnow and Santhanam [14], they were able to show that the existence of polynomial kernelizations for so-called *compositional* parameterized problems implies a collapse of the polynomial hierarchy to the third level. These are seminal results presenting the first super-linear lower bounds for kernelization and relating a statement from parameterized complexity to a hypothesis from classical complexity theory.

In Table 1 we summarize approximability and kernelization results for some well-known problems.

MIN $\mathbf{F^+\Pi_1}$ and MAX NP: Two decades ago Papadimitriou and Yannakakis [23] initiated the syntactic study of optimization problems to extend the understanding of approximability. They introduced the classes MAX NP and MAX SNP as natural variants of NP based on Fagin's [11] syntactic characterization of NP. Essentially problems are in MAX NP or MAX SNP if their optimum value can be expressed as the maximum number of tuples for which some existential, respectively quantifier-free, first-order formula holds. They showed that every problem in these two classes is approximable to within a constant factor of the optimum. Arora et al. complemented this by proving that no MAX SNP-complete problem has a polynomial-time approximation scheme, unless P=NP [2]. Contained in MAX SNP there are some well-known maximization problems, such as Max Cut, Max q-Sat, and Independent Set on graphs of bounded degree. Its superclass MAX NP also contains Max Sat amongst others.

Kolaitis and Thakur generalized the approach of examining the logical definability of optimization problems and defined further classes of minimization and maximization problems [17, 18]. Amongst others they introduced the class MIN $F^+\Pi_1$ of problems whose optimum can be expressed as the minimum weight of an assignment (i.e. number of ones) that satisfies a certain universal first-order formula. They proved that every problem in MIN $F^+\Pi_1$ is approximable to within a constant factor of the optimum. In MIN $F^+\Pi_1$ there are problems like Vertex Cover, d-Hitting Set, and Triangle Edge Deletion.

Section 2 covers the definitions of the classes MIN $F^+\Pi_1$ and MAX NP, as well as the necessary details from parameterized complexity. In Sections 3 and 4 we present polynomial kernelizations for the standard parameterizations of problems in MIN $F^+\Pi_1$ and MAX NP respectively. Section 5 summarizes our results and poses some open problems.

[2] Treewidth does not admit a polynomial kernelization unless there is a distillation algorithm for all coNP complete problems [5]. Though unlikely, this is not known to imply a collapse of the polynomial hierarchy.

2. Preliminaries

Logic and complexity classes: A (relational) vocabulary is a set σ of relation symbols, each having some fixed integer as its arity. Atomic formulas over σ are of the form $R(z_1, \dots, z_t)$ where R is a t-ary relation symbol from σ and the z_i are variables. The set of quantifier-free (relational) formulas over σ is the closure of the set of all atomic formulas under negation, conjunction, and disjunction.

Definition 2.1 (MIN $\mathrm{F}^+\Pi_1$, MAX NP)**.** A *finite structure of type* $(r_1, \dots, r_t)$ is a tuple $\mathcal{A} = (A, R_1, \dots, R_t)$ where A is a finite set and each R_i is an r_i-ary relation over A.

Let $\mathcal{Q}$ be an optimization problem on finite structures of type $(r_1, \dots, r_t)$. Let $R_1, \dots, R_t$ be relation symbols of arity $r_1, \dots, r_t$.
(a) The problem $\mathcal{Q}$ is contained in the class MIN $\mathrm{F}^+\Pi_1$ if its optimum on finite structures $\mathcal{A}$ of type $(r_1, \dots, r_t)$ can be expressed as

$$\mathrm{opt}_{\mathcal{Q}}(\mathcal{A}) = \min_S \{|S| : (\mathcal{A}, S) \models (\forall \mathbf{x} \in A^{c_x}) : \psi(\mathbf{x}, S)\},$$

where S is a single relation symbol and $\psi(\mathbf{x}, S)$ is a quantifier-free formula in conjunctive normal form over the vocabulary $\{R_1, \dots, R_t, S\}$ on variables $\{x_1, \dots, x_{c_x}\}$. Furthermore, $\psi(\mathbf{x}, S)$ is positive in S, i.e. S does not occur negated in $\psi(\mathbf{x}, S)$.
(b) The problem $\mathcal{Q}$ is contained in the class MAX NP if its optimum on finite structures $\mathcal{A}$ of type $(r_1, \dots, r_t)$ can be expressed as

$$\mathrm{opt}_{\mathcal{Q}}(\mathcal{A}) = \max_{\mathcal{S}} |\{\mathbf{x} \in A^{c_x} : (\mathcal{A}, \mathcal{S}) \models (\exists \mathbf{y} \in A^{c_y}) : \psi(\mathbf{x}, \mathbf{y}, \mathcal{S})\}|,$$

where $\mathcal{S} = (S_1, \dots, S_u)$ is a tuple of s_i-ary relation symbols S_i and $\psi(\mathbf{x}, \mathbf{y}, \mathcal{S})$ is a quantifier-free formula in disjunctive normal form over the vocabulary $\{R_1, \dots, R_t, S_1, \dots, S_u\}$ on variables $\{x_1, \dots, x_{c_x}, y_1, \dots, y_{c_y}\}$.

Remark 2.2. The definition of MAX SNP is similar to that of MAX NP but without the existential quantification of $\mathbf{y}$, i.e $\mathrm{opt}_{\mathcal{Q}}(\mathcal{A}) = \max_{\mathcal{S}} |\{\mathbf{x} : (\mathcal{A}, \mathcal{S}) \models \psi(\mathbf{x}, \mathcal{S})\}|$.

Example 2.3 (Minimum Vertex Cover)**.** Let $G = (V, E)$ be a finite structure of type (2) that represents a graph by a set V of vertices and a binary relation E over V as its edges. The optimum of Minimum Vertex Cover on structures G can be expressed as:

$$\mathrm{opt}_{VC}(G) = \min_{S \subseteq V} \{|S| : (G, S) \models (\forall (u, v) \in V^2) : (\neg E(u, v) \vee S(u) \vee S(v))\}.$$

This implies that Minimum Vertex Cover is contained in MIN $\mathrm{F}^+\Pi_1$.

Example 2.4 (Maximum Satisfiability)**.** Formulas in conjunctive normal form can be represented by finite structures $\mathcal{F} = (F, P, N)$ of type $(2, 2)$: Let F be the set of all clauses and variables, and let P and N be binary relations over F. Let $P(x, c)$ be true if and only if x is a literal of the clause c and let $N(x, c)$ be true if and only if $\neg x$ is a literal of the clause c. The optimum of Max Sat on structures $\mathcal{F}$ can be expressed as:

$$\mathrm{opt}_{MS}(\mathcal{F}) = \max_{T \subseteq F} |\{c \in F : (\mathcal{F}, T) \models (\exists x \in F) : (P(x, c) \wedge T(x)) \vee (N(x, c) \wedge \neg T(x))\}|.$$

Thus Max Sat is contained in MAX NP.

For a detailed introduction to MIN $\mathrm{F}^+\Pi_1$, MAX NP, and MAX SNP we refer the reader to [17, 18, 23]. An introduction to logic and complexity can be found in [22].

Parameterized complexity: The field of parameterized complexity, pioneered by Downey and Fellows, is a two-dimensional approach of coping with combinatorially hard problems.

Parameterized problems come with a parameterization that maps input instances to a parameter value. The time complexity of algorithms is measured with respect to the input size and the parameter. In the following we give the necessary formal definitions, namely fixed-parameter tractability, standard parameterizations, and kernelization.

Definition 2.5 (Fixed-parameter tractability)**.** A *parameterization* of Σ^* is a polynomial-time computable mapping $\kappa : \Sigma^* \to \mathbb{N}$. A *parameterized problem* over an alphabet Σ is a pair $(\mathcal{Q}, \kappa)$ consisting of a set $\mathcal{Q} \subseteq \Sigma^*$ and a parameterization κ of Σ^*.

A parameterized problem $(\mathcal{Q}, \kappa)$ is *fixed-parameter tractable* if there exists an algorithm $\mathbb{A}$, a polynomial p, and a computable function $f : \mathbb{N} \to \mathbb{N}$ such that $\mathbb{A}$ decides $x \in \mathcal{Q}$ in time $f(\kappa(x)) \cdot p(|x|)$. FPT is the class of all fixed-parameter tractable problems.

Definition 2.6 (Standard parameterization)**.** Let $\mathcal{Q}$ be a maximization (minimization) problem. The *standard parameterization* of $\mathcal{Q}$ is $p\text{-}\mathcal{Q} = (d\text{-}\mathcal{Q}, \kappa)$ where $\kappa : (\mathcal{A}, k) \mapsto k$ and $d\text{-}\mathcal{Q}$ is the language of all tuples $(\mathcal{A}, k)$ such that $\mathrm{opt}_{\mathcal{Q}}(\mathcal{A}) \geq k$ ($\mathrm{opt}_{\mathcal{Q}}(\mathcal{A}) \leq k$).

Basically $d\text{-}\mathcal{Q}$ is the decision version of $\mathcal{Q}$, asking whether the optimum is at least k (respectively at most k). The standard parameterization of $\mathcal{Q}$ is $d\text{-}\mathcal{Q}$ parameterized by k.

Definition 2.7 (Kernelization)**.** Let $(\mathcal{Q}, \kappa)$ be a parameterized problem over Σ. A polynomial-time computable function $K : \Sigma^* \to \Sigma^*$ is a *kernelization* of $(\mathcal{Q}, \kappa)$ if there is a computable function $h : \mathbb{N} \to \mathbb{N}$ such that for all $x \in \Sigma^*$ we have

$$(x \in \mathcal{Q} \Leftrightarrow K(x) \in \mathcal{Q}) \text{ and } |K(x)| \leq h(\kappa(x)).$$

We call h the *size* of the *problem kernel* $K(x)$. The kernelization K is *polynomial* if h is a polynomial. We say that $(\mathcal{Q}, \kappa)$ *admits a (polynomial) kernelization* if there exists a (polynomial) kernelization of $(\mathcal{Q}, \kappa)$.

Essentially, a kernelization is a polynomial-time data reduction that comes with a guaranteed upper bound on the size of the resulting instance in terms of the parameter.

For an introduction to parameterized complexity we refer the reader to [9, 13, 20].

Hypergraphs and sunflowers: We assume the reader to be familiar with the basic graph notation. A *hypergraph* is a tuple $\mathcal{H} = (V, E)$ consisting of a finite set V, its vertices, and a family E of subsets of V, its edges. A hypergraph has *dimension* d if each edge has cardinality at most d. A hypergraph is *d-uniform* if each edge has cardinality exactly d.

Definition 2.8 (Sunflower)**.** Let $\mathcal{H}$ be a hypergraph. A *sunflower* of cardinality r is a set $F = \{f_1, \ldots, f_r\}$ of edges of $\mathcal{H}$ such that every pair has the same intersection C, i.e. for all $1 \leq i < j \leq r$: $f_i \cap f_j = C$. The set C is called the *core* of the sunflower.

Note that any family of pairwise disjoint sets is a sunflower with core $C = \emptyset$.

Lemma 2.9 (Sunflower Lemma [10])**.** *Let $k, d \in \mathbb{N}$ and let $\mathcal{H}$ be a d-uniform hypergraph with more than $(k-1)^d \cdot d!$ edges. Then there is a sunflower of cardinality k in $\mathcal{H}$. For every fixed d there is an algorithm that computes such a sunflower in time polynomial in $|E(\mathcal{H})|$.*

Corollary 2.10 (Sunflower Corollary)**.** *The same holds for d-dimensional hypergraphs with more than $(k-1)^d \cdot d! \cdot d$ edges.*

Proof. For some $d' \in \{1, \ldots, d\}$, $\mathcal{H}$ has more than $(k-1)^d \cdot d! \geq (k-1)^{d'} \cdot d'!$ edges of cardinality d'. Let $\mathcal{H}_{d'}$ be the d'-uniform subgraph induced by the edges of cardinality d'. We apply the Sunflower Lemma on $\mathcal{H}_{d'}$ and obtain a sunflower F of cardinality k in time polynomial in $|E(\mathcal{H}_{d'})| \leq |E(\mathcal{H})|$. Clearly F is also a sunflower of $\mathcal{H}$. ■

3. Polynomial kernelization for MIN $\mathbf{F}^+\Pi_1$

We will prove that the standard parameterization of any problem in MIN $\mathrm{F}^+\Pi_1$ admits a polynomial kernelization. The class MIN $\mathrm{F}^+\Pi_1$ was introduced by Kolaitis and Thakur in a framework of syntactically defined classes of optimization problems [17]. In a follow-up paper they showed that every problem in MIN $\mathrm{F}^+\Pi_1$ is constant-factor approximable [18].

Throughout the section let $\mathcal{Q} \in$ MIN $\mathrm{F}^+\Pi_1$ be an optimization problem on finite structures of type $(r_1, \dots, r_t)$. Let $R_1, \dots, R_t$ be relation symbols of arity $r_1, \dots, r_t$ and let S be a relation symbol of arity c_S. Furthermore, let $\psi(\mathbf{x}, S)$ be a quantifier-free formula in conjunctive normal form over the vocabulary $\{R_1, \dots, R_t, S\}$ on variables $\{x_1, \dots, x_{c_x}\}$ that is positive in S such that

$$\mathrm{opt}_{\mathcal{Q}}(\mathcal{A}) = \min_{S \subseteq A^{c_S}} \{|S| : (\mathcal{A}, S) \models (\forall \mathbf{x} \in A^{c_x}) : \psi(\mathbf{x}, S)\}.$$

Let s be the maximum number of occurrences of S in any clause of $\psi(\mathbf{x}, S)$. The standard parameterization p-$\mathcal{Q}$ of $\mathcal{Q}$ is the following problem:

Input: A finite structure $\mathcal{A}$ of type $(r_1, \dots, r_t)$ and an integer k.
Parameter: k.
Task: Decide whether $\mathrm{opt}_{\mathcal{Q}}(\mathcal{A}) \leq k$.

We will see that, given an instance $(\mathcal{A}, k)$, deciding whether $\mathrm{opt}_{\mathcal{Q}}(\mathcal{A}) \leq k$ is equivalent to deciding an instance of s-HITTING SET.[3] Our kernelization will therefore make use of existing kernelization results for s-HITTING SET. The parameterized version of s-HITTING SET is defined as follows:

Input: A hypergraph $\mathcal{H} = (V, E)$ of dimension s and an integer k.
Parameter: k.
Task: Decide whether $\mathcal{H}$ has a hitting set of size at most k, i.e. $S \subseteq V$, $|S| \leq k$, such that S has a nonempty intersection with every edge of $\mathcal{H}$.

We consider the formula $\psi(\mathbf{x}, S)$ and a fixed instance $(\mathcal{A}, k)$, with $\mathcal{A} = (A, R_1, \dots, R_t)$. For every tuple $\mathbf{x} \in A^{c_x}$ we can evaluate all literals of the form $R_i(\mathbf{z})$ and $\neg R_i(\mathbf{z})$ for some $\mathbf{z} \in \{x_1, \dots, x_{c_x}\}^{r_i}$. By checking whether $\mathbf{z} \in R_i$, we obtain 1 (true) or 0 (false) for each literal. Then we delete all occurrences of 0 from the clauses and delete all clauses that contain a 1. For each $\mathbf{x}$, we obtain an equivalent formula that we denote with $\psi_{\mathbf{x}}(S)$. Each $\psi_{\mathbf{x}}(S)$ is in conjunctive normal form on literals $S(\mathbf{z})$ for some $\mathbf{z} \in \{x_1, \dots, x_{c_x}\}^{c_S}$ (no literals of the form $\neg S(\mathbf{z})$ since $\psi(\mathbf{x}, S)$ is positive in S).

Remark 3.1. For all $\mathbf{x} \in A^{c_x}$ and $S \subseteq A^{c_S}$ it holds that $(\mathcal{A}, S) \models \psi(\mathbf{x}, S)$ if and only if $(\mathcal{A}, S) \models \psi_{\mathbf{x}}(S)$. Moreover, we can compute all formulas $\psi_{\mathbf{x}}(S)$ for $\mathbf{x} \in A^{c_x}$ in polynomial time, since c_x and the length of $\psi(\mathbf{x}, S)$ are constants independent of $\mathcal{A}$.

Deriving a formula $\psi_{\mathbf{x}}(S)$ can yield empty clauses. This happens when all literals $R_i(\cdot)$, $\neg R_i(\cdot)$ in a clause are evaluated to 0 and there are no literals $S(\cdot)$. In that case, no assignment S can satisfy the formula $\psi_{\mathbf{x}}(S)$, or equivalently $\psi(\mathbf{x}, S)$. Thus $(\mathcal{A}, k)$ is a no-instance. Note that clauses of $\psi_{\mathbf{x}}(S)$ cannot contain contradicting literals since $\psi(\mathbf{x}, S)$ is positive in S.

Remark 3.2. From now on, we assume that all clauses of the formulas $\psi_{\mathbf{x}}(S)$ are nonempty.

We define a mapping Φ from finite structures $\mathcal{A}$ to hypergraphs $\mathcal{H}$. Then we show that equivalent s-HITTING SET instances can be obtained in this way.

[3] In literature the problem is often called d-HITTING SET but we will need $d = s$.

Definition 3.3. Let $\mathcal{A}$ be an instance of $\mathcal{Q}$. We define $\Phi(\mathcal{A}) := \mathcal{H}$ with $\mathcal{H} = (V, E)$. We let E be the family of all sets $e = \{\mathbf{z}_1, \dots, \mathbf{z}_p\}$ such that $(S(\mathbf{z}_1) \vee \dots \vee S(\mathbf{z}_p))$ is a clause of a $\psi_{\mathbf{x}}(S)$ for some $\mathbf{x} \in A^{c_x}$. We let V be the union of all sets $e \in E$.

Remark 3.4. The hypergraphs $\mathcal{H}$ obtained from the mapping Φ have dimension s since each $\psi_{\mathbf{x}}(S)$ has at most s literals per clause. It follows from Remark 3.1 that $\Phi(\mathcal{A})$ can be computed in polynomial time.

The following lemma establishes that $(\mathcal{A}, k)$ and $(\mathcal{H}, k) = (\Phi(\mathcal{A}), k)$ are equivalent in the sense that $(\mathcal{A}, k) \in p\text{-}\mathcal{Q}$ if and only if $(\mathcal{H}, k) \in s$-Hitting Set.

Lemma 3.5. *Let $\mathcal{A} = (A, R_1, \dots, R_t)$ be an instance of $\mathcal{Q}$ then for all $S \subseteq A^{c_S}$:*

$$(\mathcal{A}, S) \models (\forall \mathbf{x}) : \psi(\mathbf{x}, S) \text{ if and only if } S \text{ is a hitting set for } \mathcal{H} = \Phi(\mathcal{A}).$$

Proof. Let $\mathcal{H} = \Phi(\mathcal{A}) = (V, E)$ and let $S \subseteq A^{c_S}$:

$$\begin{aligned}
& (\mathcal{A}, S) \models (\forall \mathbf{x} \in A^{c_x}) : \psi(\mathbf{x}, S) \\
\Leftrightarrow\ & (\mathcal{A}, S) \models (\forall \mathbf{x} \in A^{c_x}) : \psi_{\mathbf{x}}(S) \\
\Leftrightarrow\ & (\forall \mathbf{x} \in A^{c_x}) : \text{each clause of } \psi_{\mathbf{x}}(S) \text{ has a literal } S(\mathbf{z}) \text{ for which } \mathbf{z} \in S \\
\Leftrightarrow\ & S \text{ has a nonempty intersection with every set } e \in E \\
\Leftrightarrow\ & S \text{ is a hitting set for } (V, E).
\end{aligned}$$

■

Our kernelization will consist of the following steps:

(1) Map the given instance $(\mathcal{A}, k)$ for $p\text{-}\mathcal{Q}$ to an equivalent instance $(\mathcal{H}, k) = (\Phi(\mathcal{A}), k)$ for s-Hitting Set according to Definition 3.3 and Lemma 3.5.
(2) Use a polynomial kernelization for s-Hitting Set on $(\mathcal{H}, k)$ to obtain an equivalent instance $(\mathcal{H}', k)$ with size polynomial in k.
(3) Use $(\mathcal{H}', k)$ to derive an equivalent instance $(\mathcal{A}', k)$ of $p\text{-}\mathcal{Q}$. That way we will be able to conclude that $(\mathcal{A}', k)$ is equivalent to $(\mathcal{H}, k)$ and hence also to $(\mathcal{A}, k)$.

There exist different kernelizations for s-Hitting Set: one by Flum and Grohe [13] based on the Sunflower Lemma due to Erdős and Rado [10], one by Nishimura et al. [21] via a generalization of the Nemhauser-Trotter kernelization for Vertex Cover, and a recent one by Abu-Khzam [1] based on crown decompositions. For our purposes of deriving an equivalent instance for $p\text{-}\mathcal{Q}$, these kernelizations have the drawback of shrinking sets during the reduction. This is not possible for our approach since we would need to change the formula $\psi(\mathbf{x}, S)$ to shrink the clauses. We prefer to modify Flum and Grohe's kernelization such that it uses only edge deletions.

Theorem 3.6. *There exists a polynomial kernelization of s-Hitting Set that, given an instance $(\mathcal{H}, k)$, computes an instance $(\mathcal{H}^*, k)$ such that $E(\mathcal{H}^*) \subseteq E(\mathcal{H})$, $\mathcal{H}^*$ has $O(k^s)$ edges, and the size of $(\mathcal{H}^*, k)$ is $O(k^s)$ as well.*

Proof. Let $(\mathcal{H}, k)$ be an instance of s-Hitting Set, with $\mathcal{H} = (V, E)$. If $\mathcal{H}$ contains a sunflower $F = \{f_1, \dots, f_{k+1}\}$ of cardinality $k + 1$ then every hitting set of $\mathcal{H}$ must have a nonempty intersection with the core C of F or with the $k+1$ disjoint sets $f_1 \setminus C, \dots, f_{k+1} \setminus C$. Thus every hitting set of at most k elements must have a nonempty intersection with C.

Now consider a sunflower $F = \{f_1, \dots, f_{k+1}, f_{k+2}\}$ of cardinality $k+2$ in $\mathcal{H}$ and let $\mathcal{H}' = (V, E \setminus \{f_{k+2}\})$. We show that the instances $(\mathcal{H}, k)$ and $(\mathcal{H}', k)$ are equivalent. Clearly every hitting set for $\mathcal{H}$ is also a hitting set for $\mathcal{H}'$ since $E(\mathcal{H}') \subseteq E(\mathcal{H})$. Let $S \subseteq V$ be a hitting set of size at most k for $\mathcal{H}'$. Since $F \setminus \{f_{k+2}\}$ is a sunflower of cardinality $k + 1$ in $\mathcal{H}'$, it follows that S has a nonempty intersection with its core C. Hence S has a nonempty intersection with $f_{k+2} \supseteq C$ too. Thus S is a hitting set of size at most k for $\mathcal{H}$, implying that $(\mathcal{H}, k)$ and $(\mathcal{H}', k)$ are equivalent.

We start with $\mathcal{H}^* = \mathcal{H}$ and repeat the following step while $\mathcal{H}^*$ has more than $(k+1)^s \cdot s! \cdot s$ edges. By the Sunflower Corollary we obtain a sunflower of cardinality $k+2$ in $\mathcal{H}^*$ in time polynomial in $|E(\mathcal{H}^*)|$. We delete an edge of the detected sunflower from the edge set of $\mathcal{H}^*$ (thereby reducing the cardinality of the sunflower to $k+1$). Thus, by the argument from the previous paragraph, we maintain that $(\mathcal{H}, k)$ and $(\mathcal{H}^*, k)$ are equivalent.

Furthermore $E(\mathcal{H}^*) \subseteq E(\mathcal{H})$ and $\mathcal{H}^*$ has no more than $(k+1)^s \cdot s! \cdot s \in O(k^s)$ edges. Since we delete an edge of $\mathcal{H}^*$ in each step, there are $O(|E(\mathcal{H})|)$ steps, and the total time is polynomial in $|E(\mathcal{H})|$. Deleting all isolated vertices from $\mathcal{H}^*$ yields a size of $O(s \cdot k^s) = O(k^s)$ since each edge contains at most s vertices. ∎

The following lemma proves that every s-Hitting Set instance that is "sandwiched" between two equivalent instances must be equivalent to both.

Lemma 3.7. *Let $(\mathcal{H}, k)$ be an instance of s-*Hitting Set *and let $(\mathcal{H}^*, k)$ be an equivalent instance with $E(\mathcal{H}^*) \subseteq E(\mathcal{H})$. Then for any $\mathcal{H}'$ with $E(\mathcal{H}^*) \subseteq E(\mathcal{H}') \subseteq E(\mathcal{H})$ the instance $(\mathcal{H}', k)$ is equivalent to $(\mathcal{H}, k)$ and $(\mathcal{H}^*, k)$.*

Proof. Observe that hitting sets for $\mathcal{H}$ can be projected to hitting sets for $\mathcal{H}'$ (i.e. restricted to the vertex set of $\mathcal{H}'$) since $E(\mathcal{H}') \subseteq E(\mathcal{H})$. Thus if $(\mathcal{H}, k)$ is a yes-instance then $(\mathcal{H}', k)$ is a yes-instance too. The same argument holds for $(\mathcal{H}', k)$ and $(\mathcal{H}^*, k)$. Together with the fact that $(\mathcal{H}, k)$ and $(\mathcal{H}^*, k)$ are equivalent, this proves the lemma. ∎

Now we are well equipped to prove that p-$\mathcal{Q}$ admits a polynomial kernelization.

Theorem 3.8. *Let $\mathcal{Q} \in$ MIN $\mathrm{F}^+\Pi_1$. The standard parameterization p-$\mathcal{Q}$ of $\mathcal{Q}$ admits a polynomial kernelization.*

Proof. Let $(\mathcal{A}, k)$ be an instance of p-$\mathcal{Q}$. By Lemma 3.5 we have that $(\mathcal{A}, k)$ is a yes-instance of p-$\mathcal{Q}$ if and only if $(\mathcal{H}, k) = (\Phi(\mathcal{A}), k))$ is a yes-instance of s-Hitting Set. We apply the kernelization from Theorem 3.6 to $(\mathcal{H}, k)$ and obtain an equivalent s-Hitting Set instance $(\mathcal{H}^*, k)$ such that $E(\mathcal{H}^*) \subseteq E(\mathcal{H})$ and $\mathcal{H}^*$ has $O(k^s)$ edges.

Recall that every edge of $\mathcal{H}$, say $\{\mathbf{z}_1, \ldots, \mathbf{z}_p\}$, corresponds to a clause $(S(\mathbf{z}_1) \vee \cdots \vee S(\mathbf{z}_p))$ of $\psi_{\mathbf{x}}(S)$ for some $\mathbf{x} \in A^{c_x}$. Thus for each edge $e \in E(\mathcal{H}^*) \subseteq E(\mathcal{H})$ we can select a tuple $\mathbf{x}_e$ such that e corresponds to a clause of $\psi_{\mathbf{x}_e}(S)$. Let X be the set of the selected tuples $\mathbf{x}_e$ for all edges $e \in E(\mathcal{H}^*)$. Let $A' \subseteq A$ be the set of all components of tuples $\mathbf{x}_e \in X$, ensuring that $X \subseteq A'^{c_x}$. Let R_i' be the restriction of R_i to A' and let $\mathcal{A}' = (A', R_1', \ldots, R_t')$.

Let $(\mathcal{H}', k) = (\Phi(\mathcal{A}'), k)$. By definition of Φ and by construction of $\mathcal{H}'$ we know that $E(\mathcal{H}^*) \subseteq E(\mathcal{H}') \subseteq E(\mathcal{H})$ since $X \subseteq A'^{c_x}$ and $A' \subseteq A$. Thus, by Lemma 3.7, we have that $(\mathcal{H}', k)$ is equivalent to $(\mathcal{H}, k)$. Furthermore, by Lemma 3.5, $(\mathcal{H}', k)$ is a yes-instance of s-Hitting Set if and only if $(\mathcal{A}', k)$ is a yes-instance of p-$\mathcal{Q}$. Thus $(\mathcal{A}', k)$ and $(\mathcal{A}, k)$ are equivalent instances of p-$\mathcal{Q}$.

We conclude the proof by giving an upper bound on the size of $(\mathcal{A}', k)$ that is polynomial in k. The set X contains at most $|E(\mathcal{H}^*)| \in O(k^s)$ tuples. These tuples have no more than $c_x \cdot |E(\mathcal{H}^*)|$ different components. Hence the size of A' is $O(c_x \cdot k^s) = O(k^s)$. Thus the size of $(\mathcal{A}', k)$ is $O(k^{sm})$, where m is the largest arity of a relation R_i. The values c_x, s, and m are constants that are independent of the input $(\mathcal{A}, k)$. Thus $(\mathcal{A}', k)$ is an instance equivalent to $(\mathcal{A}, k)$ with size polynomial in k. ∎

4. Polynomial kernelization for MAX NP

We prove that the standard parameterization of any problem in MAX NP admits a polynomial kernelization. The class MAX NP was introduced by Papadimitriou and Yannakakis in [23]. They showed that every problem in MAX NP is constant-factor approximable.

Throughout the section let $\mathcal{Q} \in$ MAX NP be an optimization problem on finite structures of type $(r_1, \ldots, r_t)$. Let $R_1, \ldots, R_t$ be relation symbols of arity $r_1, \ldots, r_t$ and let $\mathcal{S} = (S_1, \ldots, S_u)$ be a tuple of relation symbols of arity $s_1, \ldots, s_u$. Let $\psi(\mathbf{x}, \mathbf{y}, \mathcal{S})$ be a formula in disjunctive normal form over the vocabulary $\{R_1, \ldots, R_t, S_1, \ldots, S_u\}$ on variables $\{x_1, \ldots, x_{c_x}, y_1, \ldots, y_{c_y}\}$ such that for all finite structures $\mathcal{A}$ of type $(r_1, \ldots, r_t)$:

$$\mathrm{opt}_{\mathcal{Q}}(\mathcal{A}) = \max_{\mathcal{S}} |\{\mathbf{x} \in A^{c_x} : (\mathcal{A}, \mathcal{S}) \models (\exists \mathbf{y} \in A^{c_y}) : \psi(\mathbf{x}, \mathbf{y}, \mathcal{S})\}|.$$

Let s be the maximum number of occurrences of relations $S_1, \ldots, S_u$ in any disjunct of $\psi(\mathbf{x}, \mathbf{y}, \mathcal{S})$. The standard parameterization p-$\mathcal{Q}$ of $\mathcal{Q}$ is the following problem:

Input: A finite structure $\mathcal{A}$ of type $(r_1, \ldots, r_t)$ and an integer k.
Parameter: k.
Task: Decide whether $\mathrm{opt}_{\mathcal{Q}}(\mathcal{A}) \geq k$.

Similarly to the previous section, we consider the formula $\psi(\mathbf{x}, \mathbf{y}, \mathcal{S})$ and a fixed instance $(\mathcal{A}, k)$ with $\mathcal{A} = (A, R_1, \ldots, R_t)$. We select tuples $\mathbf{x} \in A^{c_x}$ and $\mathbf{y} \in A^{c_y}$ and evaluate all literals of the form $R_i(\mathbf{z})$ and $\neg R_i(\mathbf{z})$ for some $\mathbf{z} \in \{x_1, \ldots, x_{x_c}, y_1, \ldots, y_{c_y}\}^{r_i}$. By checking whether $\mathbf{z} \in R_i$ we obtain 1 (true) or 0 (false) for each literal. Since $\psi(\mathbf{x}, \mathbf{y}, \mathcal{S})$ is in disjunctive normal form, we delete all occurrences of 1 from the disjuncts and delete all disjuncts that contain a 0. Furthermore, we delete all disjuncts that contain contradicting literals $S_j(\mathbf{z}), \neg S_j(\mathbf{z})$ since they cannot be satisfied. We explicitly allow empty disjuncts that are satisfied by definition for the sake of simplicity (they occur when all literals in a disjunct are evaluated to 1). We obtain an equivalent formula that we denote with $\psi_{\mathbf{x},\mathbf{y}}(\mathcal{S})$.

Remark 4.1. For all $\mathbf{x}$, $\mathbf{y}$, and $\mathcal{S}$ it holds that $(\mathcal{A}, \mathcal{S}) \models \psi(\mathbf{x}, \mathbf{y}, \mathcal{S})$ iff $(\mathcal{A}, \mathcal{S}) \models \psi_{\mathbf{x},\mathbf{y}}(\mathcal{S})$. Moreover, we can compute all formulas $\psi_{\mathbf{x},\mathbf{y}}(\mathcal{S})$ for $\mathbf{x} \in A^{c_x}$, $\mathbf{y} \in A^{c_y}$ in polynomial time, since c_x, c_y, and the length of $\psi(\mathbf{x}, \mathbf{y}, \mathcal{S})$ are constants independent of $\mathcal{A}$.

Definition 4.2. Let $\mathcal{A} = (A, R_1, \ldots, R_t)$ be a finite structure of type $(r_1, \ldots, r_t)$.
(a) We define $X_{\mathcal{A}} \subseteq A^{c_x}$ as the set of all tuples $\mathbf{x}$ such that $(\exists \mathbf{y}) : \psi_{\mathbf{x},\mathbf{y}}(\mathcal{S})$ holds for some $\mathcal{S}$:

$$X_{\mathcal{A}} = \{\mathbf{x} : (\exists \mathcal{S}) : (\mathcal{A}, \mathcal{S}) \models (\exists \mathbf{y}) : \psi_{\mathbf{x},\mathbf{y}}(\mathcal{S})\}.$$

(b) For $\mathbf{x} \in A^{c_x}$ we define $Y_{\mathcal{A}}(\mathbf{x})$ as the set of all tuples $\mathbf{y}$ such that $\psi_{\mathbf{x},\mathbf{y}}(\mathcal{S})$ holds for some $\mathcal{S}$:

$$Y_{\mathcal{A}}(\mathbf{x}) = \{\mathbf{y} : (\exists \mathcal{S}) : (\mathcal{A}, \mathcal{S}) \models \psi_{\mathbf{x},\mathbf{y}}(\mathcal{S})\}.$$

Remark 4.3. The sets $X_{\mathcal{A}}$ and $Y_{\mathcal{A}}(\mathbf{x})$ can be computed in polynomial time since the number of tuples $\mathbf{x} \in A^{c_x}$ and $\mathbf{y} \in A^{c_y}$ is polynomial in the size of A and $\psi(\mathbf{x}, \mathbf{y}, \mathcal{S})$ is of constant length independent of $\mathcal{A}$.

Lemma 4.4. *Let $(\mathcal{A}, k)$ be an instance of p-$\mathcal{Q}$. If $|X_{\mathcal{A}}| \geq k \cdot 2^s$ then $\mathrm{opt}_{\mathcal{Q}}(\mathcal{A}) \geq k$, i.e. $(\mathcal{A}, k)$ is a yes-instance.*

Proof. The lemma can be concluded from the proof of the constant-factor approximability of problems in MAX NP in [23]. For each $\mathbf{x} \in X_{\mathcal{A}}$ we fix a tuple $\mathbf{y} \in Y_{\mathcal{A}}(\mathbf{x})$ such that $\psi_{\mathbf{x},\mathbf{y}}(\mathcal{S})$ is satisfiable. This yields $|X_{\mathcal{A}}|$ formulas, say $\psi_1, \ldots, \psi_{|X_{\mathcal{A}}|}$. Papadimitriou and Yannakakis showed that one can efficiently compute an assignment that satisfies at least $\sum f_i$ of these formulas, where f_i is the fraction of all assignments that satisfies ψ_i.

To see that $f_i \geq 2^{-s}$; consider such a formula ψ_i. Since ψ_i is satisfiable there exists a satisfiable disjunct. To satisfy a disjunct of at most s literals, at most s variables need to be assigned accordingly. Since the assignment to all other variables can be arbitrary this implies that $f_i \geq 2^{-s}$. Thus we have that $\sum f_i \geq |X_{\mathcal{A}}| \cdot 2^{-s}$. Therefore $|X_{\mathcal{A}}| \geq k \cdot 2^s$ implies that the assignment satisfies at least k formulas, i.e. that $\mathrm{opt}_{\mathcal{Q}}(\mathcal{A}) \geq k$. ■

Henceforth we assume that $|X_{\mathcal{A}}| < k \cdot 2^s$.

Definition 4.5. Let $(\mathcal{A}, k)$ be an instance of p-$\mathcal{Q}$ with $\mathcal{A} = (A, R_1, \ldots, R_t)$. For $\mathbf{x} \in A^{c_x}$ we define $D_{\mathcal{A}}(\mathbf{x})$ as the set of all disjuncts of $\psi_{\mathbf{x},\mathbf{y}}(\mathcal{S})$ for $\mathbf{y} \in Y_{\mathcal{A}}(\mathbf{x})$.

Definition 4.6. We define the *intersection of two disjuncts* as the conjunction of all literals that occur in both disjuncts. A *sunflower of a set of disjuncts* is a subset such that each pair of disjuncts in the subset has the same intersection (modulo permutation of the literals).

Remark 4.7. The size of each $D_{\mathcal{A}}(\mathbf{x})$ is bounded by the size of $Y_{\mathcal{A}}(\mathbf{x}) \subseteq A^{c_y}$ times the number of disjuncts of $\psi(\mathbf{x}, \mathbf{y}, \mathcal{S})$ which is a constant independent of $\mathcal{A}$. Thus the size of each $D_{\mathcal{A}}(\mathbf{x})$ is bounded by a polynomial in the input size. The definition of intersection and sunflowers among disjuncts is a direct analog that treats disjuncts as sets of literals.

Definition 4.8. A *partial assignment* is a set L of literals such that no literal is the negation of another literal in L. A formula is *satisfiable under* L if there exists an assignment that satisfies the formula and each literal in L.

Proposition 4.9. *Let $(\mathcal{A}, k)$ be an instance of p-$\mathcal{Q}$. For each $\mathbf{x} \in A^{c_x}$ there exists a set $D^*_{\mathcal{A}}(\mathbf{x}) \subseteq D_{\mathcal{A}}(\mathbf{x})$ of cardinality $O(k^s)$ such that:*

(1) *For every partial assignment L of at most sk literals, $D^*_{\mathcal{A}}(\mathbf{x})$ contains a disjunct satisfiable under L, if and only if $D_{\mathcal{A}}(\mathbf{x})$ contains a disjunct satisfiable under L.*

(2) *$D^*_{\mathcal{A}}(\mathbf{x})$ can be computed in time polynomial in $|\mathcal{A}|$.*

Proof. Let $\mathcal{A} = (A, R_1, \ldots, R_t)$ be a finite structure of type $(r_1, \ldots, r_t)$, let $\mathbf{x} \in A^{c_x}$, and let $D_{\mathcal{A}}(\mathbf{x})$ be a set of disjuncts according to Definition 4.5. From the Sunflower Corollary we can derive a polynomial-time algorithm that computes a set $D^*_{\mathcal{A}}(\mathbf{x})$ by successively shrinking sunflowers. We start by setting $D^*_{\mathcal{A}}(\mathbf{x}) = D_{\mathcal{A}}(\mathbf{x})$ and apply the following step while the cardinality of $D^*_{\mathcal{A}}(\mathbf{x})$ is greater than $(sk+1)^s \cdot s! \cdot s$.

We compute a sunflower of cardinality $sk+2$, say $F = \{f_1, \ldots, f_{sk+2}\}$, in time polynomial in $|D^*_{\mathcal{A}}(\mathbf{x})|$ (Sunflower Corollary). We delete a disjunct of F, say f_{sk+2}, from $D^*_{\mathcal{A}}(\mathbf{x})$. Let O and P be copies of $D^*_{\mathcal{A}}$ before respectively after deleting f_{sk+2}. Observe that $F' = F \setminus \{f_{sk+2}\}$ is a sunflower of cardinality $sk+1$ in P. Let L be a partial assignment of at most sk literals and assume that no disjunct in P is satisfiable under L. This means that for each disjunct of P there is a literal in L that contradicts it, i.e. a literal that is the negation of a literal in the disjunct. We focus on the sunflower F' in P. There must be a literal in L, say l, that contradicts at least two disjuncts of F', say f and f', since $|F'| = sk+1$ and $|L| \leq sk$. Therefore l is the negation of a literal in the intersection of f and f', i.e. the core of F'. Thus l contradicts also f_{sk+2} and we conclude that no disjunct in $O = P \cup \{f_{sk+2}\}$ is satisfiable under the partial assignment L. The reverse argument holds since all disjuncts of P are contained in O. Thus each step maintains the desired property (1).

At the end $D^*_{\mathcal{A}}(x)$ contains no more than $(sk+1)^s \cdot s! \cdot s \in O(k^s)$ disjuncts. For each $\mathbf{x}$ this takes time polynomial in the size of the input since the cardinality of $D_{\mathcal{A}}(\mathbf{x})$ is bounded by a polynomial in the input size and a disjunct is deleted in each step. ■

Lemma 4.10. *Let $D'_{\mathcal{A}}(\mathbf{x})$ be a subset of $D_{\mathcal{A}}(\mathbf{x})$ such that $D^*_{\mathcal{A}}(\mathbf{x}) \subseteq D'_{\mathcal{A}}(\mathbf{x}) \subseteq D_{\mathcal{A}}(\mathbf{x})$. For any partial assignment L of at most sk literals it holds that $D_{\mathcal{A}}(\mathbf{x})$ contains a disjunct satisfiable under L if and only if $D'_{\mathcal{A}}(\mathbf{x})$ contains a disjunct satisfiable under L.*

Proof. Let L be a partial assignment of at most sk literals. If $D_{\mathcal{A}}(\mathbf{x})$ contains a disjunct satisfiable under L, then, by Proposition 4.9, this holds also for $D^*_{\mathcal{A}}(\mathbf{x})$. For $D^*_{\mathcal{A}}(\mathbf{x})$ and $D'_{\mathcal{A}}$ this holds since $D^*_{\mathcal{A}}(\mathbf{x}) \subseteq D'_{\mathcal{A}}(\mathbf{x})$. The same is true for $D'_{\mathcal{A}}(\mathbf{x})$ and $D_{\mathcal{A}}(\mathbf{x})$. ■

Theorem 4.11. *Let $\mathcal{Q} \in$ MAX NP. The standard parameterization p-$\mathcal{Q}$ of $\mathcal{Q}$ admits a polynomial kernelization.*

Proof. The proof is organized in three parts. First, given an instance $(\mathcal{A}, k)$ of p-$\mathcal{Q}$, we construct an instance $(\mathcal{A}', k)$ of p-$\mathcal{Q}$ in time polynomial in the size of $(\mathcal{A}, k)$. In the second part, we prove that $(\mathcal{A}, k)$ and $(\mathcal{A}', k)$ are equivalent. In the third part, we conclude the proof by showing that the size of $(\mathcal{A}', k)$ is bounded by a polynomial in k.

(I.) Let $(\mathcal{A}, k)$ be an instance of p-$\mathcal{Q}$. We use the sets $D_{\mathcal{A}}(\mathbf{x})$ and $D^*_{\mathcal{A}}(\mathbf{x})$ according to Definition 4.5 and Proposition 4.9. Recall that $D_{\mathcal{A}}(\mathbf{x})$ is the set of all disjuncts of $\psi_{\mathbf{x},\mathbf{y}}(\mathcal{S})$ for $\mathbf{y} \in Y_{\mathcal{A}}(\mathbf{x})$. Thus, for each disjunct $d \in D^*_{\mathcal{A}}(\mathbf{x}) \subseteq D_{\mathcal{A}}(\mathbf{x})$, we can select a $\mathbf{y}_d \in Y_{\mathcal{A}}(\mathbf{x})$ such that d is a disjunct of $\psi_{\mathbf{x},\mathbf{y}_d}(\mathcal{S})$. Let $Y'_{\mathcal{A}}(\mathbf{x}) \subseteq Y_{\mathcal{A}}(\mathbf{x})$ be the set of these selected tuples $\mathbf{y}_d$. Let $D'_{\mathcal{A}}(\mathbf{x})$ be the set of all disjuncts of $\psi_{\mathbf{x},\mathbf{y}}(\mathcal{S})$ for $\mathbf{y} \in Y'_{\mathcal{A}}(\mathbf{x})$. Since $D^*_{\mathcal{A}}(\mathbf{x})$ contains some disjuncts of $\psi_{\mathbf{x},\mathbf{y}}(\mathcal{S})$ for $\mathbf{y} \in Y'_{\mathcal{A}}(\mathbf{x})$ and $D_{\mathcal{A}}(\mathbf{x})$ contains all disjuncts of $\psi_{\mathbf{x},\mathbf{y}}(\mathcal{S})$ for $\mathbf{y} \in Y_{\mathcal{A}}(\mathbf{x}) \supseteq Y'_{\mathcal{A}}(\mathbf{x})$, we have that $D^*_{\mathcal{A}}(\mathbf{x}) \subseteq D'_{\mathcal{A}}(\mathbf{x}) \subseteq D_{\mathcal{A}}(\mathbf{x})$.

For each $\mathbf{x}$ this takes time $O(|D^*_{\mathcal{A}}(\mathbf{x})| \cdot |Y^*_{\mathcal{A}}(\mathbf{x})|) \subseteq O(k^s \cdot |A|^{c_y})$. Computing $Y'_{\mathcal{A}}(\mathbf{x})$ for all $\mathbf{x} \in A^{c_x}$ takes time $O(|A|^{c_x} \cdot k^s \cdot |A|^{c_y})$, i.e. time polynomial in the size of $(\mathcal{A}, k)$ since k is never larger than $|A|^{c_x}$.[4]

Let $A' \subseteq A$ be the set of all components of $\mathbf{x} \in X_{\mathcal{A}}$ and $\mathbf{y} \in Y'_{\mathcal{A}}(\mathbf{x})$ for all $\mathbf{x} \in X_{\mathcal{A}}$. This ensures that $X_{\mathcal{A}} \subseteq (A')^{c_x}$ and $Y'_{\mathcal{A}}(\mathbf{x}) \subseteq (A')^{c_y}$ for all $\mathbf{x} \in X_{\mathcal{A}}$. Let R'_i be the restriction of R_i to A' and let $\mathcal{A}' = (A', R'_1, \ldots, R'_t)$.

(II.) We will now prove that $\mathrm{opt}_{\mathcal{Q}}(\mathcal{A}) \geq k$ if and only if $\mathrm{opt}_{\mathcal{Q}}(\mathcal{A}') \geq k$, i.e. that $(\mathcal{A}, k)$ and $(\mathcal{A}', k)$ are equivalent. Assume that $\mathrm{opt}_{\mathcal{Q}}(\mathcal{A}) \geq k$ and let $\mathcal{S} = (S_1, \ldots, S_u)$ such that $|\{\mathbf{x} : (\mathcal{A}, \mathcal{S}) \models (\exists \mathbf{y}) : \psi(\mathbf{x}, \mathbf{y}, \mathcal{S})\}| \geq k$. This implies that there must exist tuples $\mathbf{x}_1, \ldots, \mathbf{x}_k \in A^{c_x}$ and $\mathbf{y}_1, \ldots, \mathbf{y}_k \in A^{c_y}$ such that $\mathcal{S}$ satisfies $\psi_{\mathbf{x}_i,\mathbf{y}_i}(\mathcal{S})$ for $i = 1, \ldots, k$. Thus $\mathcal{S}$ must satisfy at least one disjunct in each $\psi_{\mathbf{x}_i,\mathbf{y}_i}(\mathcal{S})$ since these formulas are in disjunctive normal form. Accordingly let $d_1, \ldots, d_k$ be disjuncts such that $\mathcal{S}$ satisfies the disjunct d_i in $\psi_{\mathbf{x}_i,\mathbf{y}_i}(\mathcal{S})$ for $i = 1, \ldots, k$. We show that there exists $\mathcal{S}'$ such that:

$$|\{\mathbf{x} : (\mathcal{A}', \mathcal{S}') \models (\exists \mathbf{y}) : \psi(\mathbf{x}, \mathbf{y}, \mathcal{S}')\}| \geq k.$$

For $p = 1, \ldots, k$ we apply the following step: If $\mathbf{y}_p \in Y'_{\mathcal{A}}(\mathbf{x}_p)$ then do nothing. Otherwise consider the partial assignment L consisting of the at most sk literals of the disjuncts $d_1, \ldots, d_k$. The set $D_{\mathcal{A}}(\mathbf{x}_p)$ contains a disjunct that is satisfiable under L, namely d_p. By Lemma 4.10, it follows that $D'_{\mathcal{A}}(\mathbf{x}_p)$ also contains a disjunct satisfiable under L, say d'_p. Let $\mathbf{y}'_p \in Y'_{\mathcal{A}}(\mathbf{x}_p)$ such that d'_p is a disjunct of $\psi_{\mathbf{x}_p,\mathbf{y}'_p}(\mathcal{S})$. Such a $\mathbf{y}'_p$ can be found by selection of $D'_{\mathcal{A}}(\mathbf{x}_p)$. Change $\mathcal{S}$ in the following way to satisfy the disjunct d'_p. For each literal of d'_p of the form $S_i(\mathbf{z})$ add $\mathbf{z}$ to the relation S_i. Similarly for each literal of the form $\neg S_i(\mathbf{z})$

[4] That is, $(\mathcal{A}, k)$ is a no-instance if $k > |A|^{c_x}$ since k exceeds the number of tuples $\mathbf{x} \in A^{c_x}$.

remove $\mathbf{z}$ from S_i. This does not change the fact that $\mathcal{S}$ satisfies the disjunct d_i in $\psi_{\mathbf{x}_i,\mathbf{y}_i}(\mathcal{S})$ for $i = 1, \dots, k$ since, by selection, d'_p is satisfiable under L. Then we replace $\mathbf{y}_p$ by $\mathbf{y}'_p$ and d_p by d'_p. Thus we maintain that $\mathcal{S}$ satisfies d_i in $\psi_{\mathbf{x}_i,\mathbf{y}_i}(\mathcal{S})$ for $i = 1, \dots, k$.

After these steps we obtain $\mathcal{S}$ as well as tuples $\mathbf{x}_1, \dots, \mathbf{x}_k$, $\mathbf{y}_1, \dots, \mathbf{y}_k$ with $\mathbf{y}_i \in Y'_{\mathcal{A}}(\mathbf{x}_i)$, and disjuncts $d_1, \dots, d_k$ such that $\mathcal{S}$ satisfies d_i in $\psi_{\mathbf{x}_i,\mathbf{y}_i}(\mathcal{S})$ for $i = 1, \dots, k$. Let $\mathcal{S}'$ be the restriction of $\mathcal{S}$ to A'. Then we have that $(\mathcal{A}', \mathcal{S}') \models \psi_{\mathbf{x}_i,\mathbf{y}_i}(\mathcal{S}')$ for $i = 1, \dots, k$ since A' is defined to contain the components of tuples $\mathbf{x} \in X_{\mathcal{A}}$ and of all tuples $\mathbf{y} \in Y'_{\mathcal{A}}(\mathbf{x})$ for $\mathbf{x} \in X_{\mathcal{A}}$. Hence $\mathbf{x}_i \in \{\mathbf{x} : (\mathcal{A}', \mathcal{S}') \models (\exists \mathbf{y}) : \psi(\mathbf{x}, \mathbf{y}, \mathcal{S}')\}$ for $i = 1, \dots, k$. Thus $\mathrm{opt}_{\mathcal{Q}}(\mathcal{A}') \geq k$.

For the reverse direction assume that $\mathrm{opt}_{\mathcal{Q}}(\mathcal{A}') \geq k$. Since $A' \subseteq A$ it follows that

$$\{\mathbf{x} : (\mathcal{A}', \mathcal{S}') \models (\exists \mathbf{y}) : \psi(\mathbf{x}, \mathbf{y}, \mathcal{S}')\} \subseteq \{\mathbf{x} : (\mathcal{A}, \mathcal{S}') \models (\exists \mathbf{y}) : \psi(\mathbf{x}, \mathbf{y}, \mathcal{S}')\}.$$

Thus $|\{\mathbf{x} : (\mathcal{A}, \mathcal{S}') \models (\exists \mathbf{y}) : \psi(\mathbf{x}, \mathbf{y}, \mathcal{S}')\}| \geq k$, implying that $\mathrm{opt}_{\mathcal{Q}}(\mathcal{A}) \geq k$. Therefore $\mathrm{opt}_{\mathcal{Q}}(\mathcal{A}) \geq k$ if and only if $\mathrm{opt}_{\mathcal{Q}}(\mathcal{A}') \geq k$. Hence $(\mathcal{A}, k)$ and $(\mathcal{A}', k)$ are equivalent instances of p-$\mathcal{Q}$.

(III.) We conclude the proof by providing an upper bound on the size of $(\mathcal{A}', k)$ that is polynomial in k. For the sets $Y'_{\mathcal{A}}(\mathbf{x})$ we selected one tuple $\mathbf{y}$ for each disjunct in $D^*_{\mathcal{A}}(\mathbf{x})$. Thus $|Y'_{\mathcal{A}}(\mathbf{x})| \leq |D^*(\mathbf{x})| \in O(k^s)$ for all $\mathbf{x} \in X_{\mathcal{A}}$. The set A' contains the components of tuples $\mathbf{x} \in X_{\mathcal{A}}$ and of all tuples $\mathbf{y} \in Y'_{\mathcal{A}}(\mathbf{x})$ for $\mathbf{x} \in X_{\mathcal{A}}$. Thus

$$\begin{aligned} |A'| &\leq c_x \cdot |X_{\mathcal{A}}| + c_y \cdot \textstyle\sum_{\mathbf{x} \in X_{\mathcal{A}}} |Y'_{\mathcal{A}}(\mathbf{x})| \\ &\leq c_x \cdot |X_{\mathcal{A}}| + c_y \cdot |X_{\mathcal{A}}| \cdot O(k^s) \\ &< c_x \cdot k \cdot 2^s + c_y \cdot k \cdot 2^s \cdot O(k^s) = O(k^{s+1}). \end{aligned}$$

For each relation R'_i we have $|R'_i| \leq |A'|^{r_i} \in O(k^{(s+1)r_i})$. Thus the size of $(\mathcal{A}', k)$ is bounded by $O(k^{(s+1)m})$, where m is the largest arity of a relation R_i. ■

Remark 4.12. For MAX SNP one can prove a stronger result that essentially relies on Lemma 4.4. That way one obtains bounds for the sizes of A' and $(\mathcal{A}', k)$ of $O(k)$ and $O(k^m)$ respectively.

5. Conclusion

We have constructively established that the standard parameterizations of problems in MIN $\mathrm{F}^+\Pi_1$ and MAX NP admit polynomial kernelizations. Thus a strong relation between constant-factor approximability and polynomial kernelizability has been showed for two large classes of problems. It remains an open problem to give a more general result that covers all known examples (e.g. Feedback Vertex Set). It might be profitable to consider closures of MAX SNP under reductions that preserve constant-factor approximability. Khanna et al. [16] proved that APX and APX-PB are the closures of MAX SNP under PTAS-preserving reductions and E-reductions, respectively. Since both classes contain Bin Packing which does not admit a polynomial kernelization, this leads to the question whether polynomial kernelizability or fixed-parameter tractability are maintained under restricted versions of these reductions.

Acknowledgement

We would like to thank Iyad Kanj for pointing out Bin Packing as an example that is constant-factor approximable but does not admit a polynomial kernelization.

References

[1] Faisal N. Abu-Khzam. Kernelization algorithms for d-hitting set problems. In Frank K. H. A. Dehne, Jörg-Rüdiger Sack, and Norbert Zeh, editors, *WADS*, volume 4619 of *LNCS*, pages 434–445. Springer, 2007.

[2] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof verification and the hardness of approximation problems. *J. ACM*, 45(3):501–555, 1998.

[3] Vineet Bafna, Piotr Berman, and Toshihiro Fujito. A 2-approximation algorithm for the undirected feedback vertex set problem. *SIAM J. Discrete Math.*, 12(3):289–297, 1999.

[4] Hans L. Bodlaender. A cubic kernel for feedback vertex set. In Wolfgang Thomas and Pascal Weil, editors, *STACS*, volume 4393 of *LNCS*, pages 320–331. Springer, 2007.

[5] Hans L. Bodlaender, Rodney G. Downey, Michael R. Fellows, and Danny Hermelin. On problems without polynomial kernels (extended abstract). In Luca Aceto, Ivan Damgård, Leslie Ann Goldberg, Magnús M. Halldórsson, Anna Ingólfsdóttir, and Igor Walukiewicz, editors, *ICALP (1)*, volume 5125 of *LNCS*, pages 563–574. Springer, 2008.

[6] Liming Cai and Jianer Chen. On fixed-parameter tractability and approximability of NP optimization problems. *J. Comput. Syst. Sci.*, 54(3):465–474, 1997.

[7] Liming Cai and Xiuzhen Huang. Fixed-parameter approximation: Conceptual framework and approximability results. In Hans L. Bodlaender and Michael A. Langston, editors, *IWPEC*, volume 4169 of *Lecture Notes in Computer Science*, pages 96–108. Springer, 2006.

[8] Jianer Chen, Iyad A. Kanj, and Weijia Jia. Vertex cover: Further observations and further improvements. *J. Algorithms*, 41(2):280–301, 2001.

[9] Rod G. Downey and M. R. Fellows. *Parameterized Complexity (Monographs in Computer Science)*. Springer, November 1998.

[10] Paul Erdős and Richard Rado. Intersection theorems for systems of sets. *J. London Math. Soc.*, 35:85–90, 1960.

[11] Ronald Fagin. Generalized first-order spectra and polynomial-time recognizable sets. In *Complexity of Computation*, volume 7 of *SIAM-AMS Proceedings*, pages 43–73, SIAM, Philadelphia, PA, 1974.

[12] Uriel Feige, MohammadTaghi Hajiaghayi, and James R. Lee. Improved approximation algorithms for minimum weight vertex separators. *SIAM J. Comput.*, 38(2):629–657, 2008.

[13] J. Flum and M. Grohe. *Parameterized Complexity Theory (Texts in Theoretical Computer Science. An EATCS Series)*. Springer, March 2006.

[14] Lance Fortnow and Rahul Santhanam. Infeasibility of instance compression and succinct PCPs for NP. In Richard E. Ladner and Cynthia Dwork, editors, *STOC*, pages 133–142. ACM, 2008.

[15] Eran Halperin. Improved approximation algorithms for the vertex cover problem in graphs and hypergraphs. *SIAM J. Comput.*, 31(5):1608–1623, 2002.

[16] Sanjeev Khanna, Rajeev Motwani, Madhu Sudan, and Umesh V. Vazirani. On syntactic versus computational views of approximability. *SIAM J. Comput.*, 28(1):164–191, 1998.

[17] Phokion G. Kolaitis and Madhukar N. Thakur. Logical definability of NP optimization problems. *Inf. Comput.*, 115(2):321–353, 1994.

[18] Phokion G. Kolaitis and Madhukar N. Thakur. Approximation properties of NP minimization classes. *J. Comput. Syst. Sci.*, 50(3):391–411, 1995.

[19] Assaf Natanzon, Ron Shamir, and Roded Sharan. A polynomial approximation algorithm for the minimum fill-in problem. *SIAM J. Comput.*, 30(4):1067–1079, 2000.

[20] Rolf Niedermeier. *Invitation to Fixed Parameter Algorithms (Oxford Lecture Series in Mathematics and Its Applications)*. Oxford University Press, USA, March 2006.

[21] Naomi Nishimura, Prabhakar Ragde, and Dimitrios M. Thilikos. Smaller kernels for hitting set problems of constant arity. In Rodney G. Downey, Michael R. Fellows, and Frank K. H. A. Dehne, editors, *IWPEC*, volume 3162 of *LNCS*, pages 121–126. Springer, 2004.

[22] Christos H. Papadimitriou. *Computational Complexity*. Addison Wesley, November 1993.

[23] Christos H. Papadimitriou and Mihalis Yannakakis. Optimization, approximation, and complexity classes. *J. Comput. Syst. Sci.*, 43(3):425–440, 1991.